BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

WORK INTEGRATED LEARNING PROGRAMMES

COURSE HANDOUT

Part A: Content Design

Course Title	Information Retrieval
Course No(s)	AIML ZG537
Credit Units	4
Credit model	3-1-0
Course Author	Dr. Chetana Gavankar
Instructor	V. MAHESWARI (LEAD), KAMLENDRA SINGH SHEKHAWAT
Version No	2.0
Date	20-02-2025

Course Description:
Organization, representation, and access to information; categorization, indexing, and content analysis; data structures for unstructured data; design and maintenance of such data structures, indexing and indexes, retrieval and classification schemes; use of codes, formats, and standards; analysis, construction and evaluation of search and navigation techniques; search engines and how they relate to the above. Multimedia data and their representation and search.

Course Objectives

No	Course Objective
CO1	To understand the structure and organization of various components of an IR system
CO2	To understand information representation models, term scoring mechanisms, etc. in the complete search system
CO3	To understand the architecture of search engines, crawlers, and the web search
CO4	To understand cross-lingual retrieval and multimedia information retrieval

Text Book(s)

T1	C. D. Manning, P. Raghavan and H. Schutze. Introduction to Information Retrieval, Cambridge University Press, 2008. http://nlp.stanford.edu/IR-book/

Reference Book(s) & other resources

R1	Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. http://people.ischool.berkeley.edu/~hearst/irbook/
R2	Ricci, F.; Rokach, L.; Shapira, B.; Kantor, P.B. (Eds.), Recommender Systems Handbook. 1st Edition., 2011, 845 p. 20 illus., Hardcover, ISBN: 978-0-387-85819-7
R3	Cross-Language Information Retrieval by Jian-Yun Nie Morgan & Claypool Publisher series 2010
R4	Multimedia Information Retrieval by Stefan M. Rüger Morgan & Claypool Publisher series 2010.
R5	Information Retrieval: Implementing and Evaluating Search Engines by S. Buttcher, C. Clarke and G. Cormack, MIT Press, 2010.
R6	Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by B. Liu, Springer, Second Edition, 2011.

Learning Outcomes:

No	Learning Outcomes
LO1	Students will gain an understanding of an information retrieval system as a whole and about its components.
LO2	Students will have knowledge about the design issues and their solutions of different types of models including Boolean, vector space, etc.
LO3	Students will have a detailed understanding of text indexing, mining, weighting schemes, etc.
LO4	Students will acquire knowledge about cross-lingual and multimedia information retrieval.
LO5	With the acquired knowledge students will be able to design and build different kinds of information retrieval systems.

Modular Content Structure

1. Introduction
   - 1.1. Information Retrieval
   - 1.2. Basic Search Model
2. Basic Information Retrieval Concepts
   - 2.1. Boolean Retrieval
   - 2.2. Dictionaries and Tolerant Retrieval
   - 2.3. Index Construction and Compression
3. Vector Space Model
   - 3.1. Scoring, Term Weighting
   - 3.2. The Vector Space Model for Scoring
4. Text Mining
   - 4.1. Text Classification
   - 4.2. Vector Space Classification
   - 4.3. Text Clustering
5. Evaluation in Information Retrieval
   - 5.1. Evaluation in unranked retrieval sets
   - 5.2. Evaluation in ranked retrieval sets
6. Web Search
   - 6.1. Web Search Basics
   - 6.2. Web Crawlers and Indexes
   - 6.3. Link Analysis
7. Cross-Lingual Retrieval
   - 7.1. Language Problems in IR
   - 7.2. Approaches for CLIR
   - 7.3. Neural
8. Multimedia Information Retrieval
   - 8.1. Multimedia Search Technologies
   - 8.2. Content-Based Retrieval
9. Recommender Systems
   - 9.1. Collaborative and Content-Based RS
10. Neural IR
    - 10.1 Introduction to Deep neural network
    - 10.2 Deep neural network for IR

Part B: Contact Session Plan

Academic Term	Mtech DSE June 2021
Course Title	Information Retrieval
Course No	AIML/DSE CZG537
Lead Instructor	Dr. Chetana Gavankar

Contact Session	List of Topic Title (from content structure in Part A)	Topic # (from content structure in Part A)	Text/Ref Book/external resource
1	● Introduction o Information Vs Data Retrieval o Basic Concepts o The retrieval process o Taxonomy of IR o Classic IR and Alternative models	1.1, 1.2	R1 Ch1, Ch2
2	● Boolean Retrieval o Inverted index o Processing Boolean queries o Boolean Vs Ranked retrieval o Term vocabulary and postings lists o Phrase queries	2.1	T1 Ch 1, 2
3	● Dictionary and Tolerant Retrieval o Search Structures for dictionaries o Wildcard queries o Phonetic Correction	2.2	T1 Ch3
4	● Index Construction and Compression o Blocked sort-based Indexing o Single pass in-memory indexing o Distributed and dynamic indexing	2.3	T1 Ch4,5

	○ Dictionary comparison ○ Postings file compression		
5	● Vector Space Model ○ Term frequency and weighting ○ The vector space model for scoring ○ Tf-idf functions	3.1, 3.2	T1 Ch6
6	● Classification & Clustering for IR ○ Feature Selection ○ Vector space classification ○ Document Representation ○ Rocchio classification ○ Evaluating Classification	4.1	T1 Ch13, 14 T1 Ch16, 17
7	● Evaluation in Information Retrieval ○ Evaluation in unranked retrieval sets ○ Evaluation in ranked retrieval sets		T1 Chp 8
8	● Review		
9	● Web Search ○ Web characteristics ○ The search user experience ○ Index size and estimation	5.1	T1 Ch19
10	● Web Crawling and Indexes ○ Crawling ○ Crawler Architecture ○ Distributed Indexes	5.2	T1 Ch20
11	● Link Analysis ○ The web as a graph ○ Google’s page rank ○ Hub and Authorities (HITS)	5.3	T1 Ch21
12	● Cross Lingual IR (CLIR) ○ Language problems in IR ○ Translation Approaches	6.1, 6.2	R3 Ch2

	o Handling Many Languages o Resources for CLIR		
13	● Multimedia IR o Basic Multimedia search technologies o Content Based Retrieval	7.1,7.2	R4 Ch2,3
14	● Recommender System o Collaborative recommendation o Content based recommendation o Other type & hybrid recommendations	8.1	R2 Ch1-5
15	NEURAL IR 15.1.Introduction to Deep neural networks 15.1.1Input text representation 15.1.2 Standard architectures 15.2 Deep neural networks for IR 15.2.1 Document autoencoders 15.2.2 Siamese networks 15.2.3 Interaction-based network 15.2.4 Lexical and semantic matching 15.2.5 Matching with multiple document fields	6&7	An Introduction to Neural Information Retrieval by Bhaskar Mitra, Nick Craswell Link: https://www.microsoft.com/en-us/research/uploads/prod/2017/06/fntir2018-neuralir-mitra.pdf
16	● Review		

Work integration: Detailed plan

No	Activity description

	(Examples are given below)
1	Apply Domain modelling concept to the work you are doing in the workplace
2	Present the architecture of the software you are working on
3	Analyze the test plan of the software project you are working on and identify areas where it can be further improved
4	Seminar/talk by Project manager in the company on a topic of relevance to the course

Evaluation Scheme:

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

Evaluation Component	Name (Quiz, Lab, Project, Mid-term exam, End semester exam, etc.)	Type (Open book, Closed book, Online, etc.)	Weight	Duration	Day, Date, Session, Time
EC – 1*	Quiz 1	Online	5%	2days	August 10-20, 2026
	Quiz2	Online	5%	2days	August 27-7 Sept 2026
	Assignment	Online	20 %	10 days	29th Oct to 11th Nov 2026
EC - 2	Mid-Semester Test	Closed Book	30%	2 hours	20/09/2026 (FN)
EC - 3	Comprehensive Exam	Open Book	40%	2 ½ Hours	06/12/2026 (FN)

EC1* (20% - 30%): Quiz (optional): 5-10 %, Lab Assignment/Assignment: 20% - 30%

Syllabus for Mid-Semester Test (Closed Book): Topics in Contact session: 1 to 8

Syllabus for Comprehensive Exam (Open Book): All topics

Important Links and Information:

eLearn Portal: https://elearn.bits-pilani.ac.in

Students must visit the eLearn portal regularly and stay updated with the latest announcements and deadlines.

Contact Sessions: Students should attend the online lectures as per the schedule provided on the eLearn portal.

Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Students will attempt them through the course pages on the eLearn portal. Announcements will be made on the portal in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted.
3. For Open Book exams: “open book” means text/ reference books (publisher copy only) and does not include any other learning material. No other learning material will be permitted during the open book examinations. For Detailed Guidelines refer to the attached document.
   EC3 Guidelines
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam, which will be made available on the eLearn portal. The Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.
5. It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the online lectures, and take all the prescribed evaluation components such as Assignments/Quizzes, Mid-Semester Tests and Comprehensive Exams according to the evaluation scheme provided in the handout.